

1. “List of all participants’ names, affiliations, and mailing addresses.” Check for accuracy.

Organizer: Brian W. Junker
Department of Statistics
Carnegie Mellon University
Pittsburgh PA 15213
brian@stat.cmu.edu

Moderator: Kenneth R. Koedinger
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh PA 15213
koedinger@cmu.edu

Presenters: Jean-Claude Falmagne
Dept. of Cognitive Sciences
University of California – Irvine
Irvine CA 92697
jcf@uci.edu

Valerie J. Shute
Research Applications and Development
Educational Testing Service
Princeton, NJ 08541
vshute@ets.org

Neil T. Heffernan
Computer Science Department
Worcester Polytechnic Institute
Worcester MA 01060
nth@wpi.edu

Brian W. Junker
Department of Statistics
Carnegie Mellon University
Pittsburgh PA 15213
brian@stat.cmu.edu

Discussant: Mark Wilson
Graduate School of Education
University of California – Berkeley
Berkeley CA 94720
MarkW@berkeley.edu

2. “500-word summary of the topic and organizational structure of the session. The summary should make clear the distinctions between the positions taken by each presenter. In addition, type or electronically insert 150-word summaries of each participant’s paper in its respective text box.”

Symposium Title: On-demand learning-embedded benchmark assessment using classroom-accessible technology

The movement towards high stakes testing promises to encourage rigor and accountability in middle school mathematics, but there is a danger that a too-narrow focus on testing will take time and attention away from instruction (e.g., Olson, 2005). In trying to use assessment to guide instruction, how can teachers be sure that ongoing assessment will improve instruction enough to justify the cost of lost instructional time?

Two related, technology-aided developments may help resolve this dilemma: First, adaptive testing based on careful mapping of a domain, such as eighth grade mathematics, promises more efficient, detailed assessment. Second, embedding assessment into learning experiences reduces the cost of lost instructional time while providing worthwhile benchmarks of student progress. Since these developments require only modest computer and internet resources, they are widely accessible in classrooms students' homes, and elsewhere. In this symposium we will explore and compare three approaches combining these developments.

ALEKS (Assessment and LEarning in Knowledge Spaces) is based on the theory of Knowledge Spaces and their specialization, Learning Spaces (Doignon & Falgagne, 1999; Cosyn & Uzun, 2005). The key idea is to partition a set of question or problem types into subsets called *knowledge states*; a student is characterized by a knowledge state if he/she can succeed at all the question types in that state, and none of the question types in its complement. Learning spaces add partial ordering and completeness conditions that facilitate determining a student's knowledge state, identifying problem types that the student is ready to learn next, etc.

The *Adaptive Content for Evidence-based Diagnosis* (ACED; Shute, Graf & Hansen, 2006) system adapts Evidence-Centered Design (ECD; Mislevy, Steinberg & Almond, 1999) to an interactive online assessment and learning system. ECD focuses on the rigorous, interlocking development of student, evidence and task models underlying assessment. In contrast to Knowledge Space theory, the ECD student model is conceived of in terms of skills or other latent variables characterizing student proficiency. The latent attributes in the student model are linked to task performance evidence, often through statistical models such as a Bayes Nets, Item Response Theory models, etc. ACED also provides a testbed for exploring alternative modes of delivery of eight-grade mathematics content to both sighted and visually-impaired students.

A hybrid approach is instantiated in the *Assistments* system (e.g., Feng, Heffernan & Koedinger, in press), which *assists* student learning while *assessing* progress toward end-of-year exams in eighth-grade mathematics. Like Knowledge Space assessment, the Assistments approach is task-centric (development of main questions and scaffolding materials starts from released state exam items); but like ECD, Assistments directly attribute individual differences to unobservable skills or other latent variables. Unlike ECD's single-student-model approach, multiple student models are envisioned for Assistments: one student model may be appropriate for predicting performance on end-of-year accountability exams, another for providing feedback to teachers on what to teach next, etc.

We will not recommend a single best system for online classroom-accessible learning-embedded benchmark assessment, but rather contrast tools for the different goals of assessing knowledge, evaluating cognitive skills, and measuring aptitudes.

Bibliography

- Cosyn, E. and Uzun, H. (2005). Axioms for Learning Spaces. *Journal of Mathematical Psychology*. To be submitted.
- Doignon, J.-P. and Falmagne, J.-Cl. (1999). *Knowledge Spaces*. Springer, Berlin.
- Feng, M., Heffernan, N.T. and Koedinger, K.R. (in press). Addressing the Testing Challenge with a Web-Based E-Assessment System that Tutors as it Assesses. *Proceedings of the Fifteenth International World Wide Web Conference*. ACM Press (Anticipated): New York, NY. 2006. Downloaded August 2006 from http://web.cs.wpi.edu/Research/trg/public/project/papers/www/feng_et al.pdf
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design*. Downloaded August 2006 from http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html
- Olson, L. (2005). Benchmark assessments offer regular checkups on student achievement. *Education Week*, November 30, 2005. Downloaded April 2006 from <http://www.edweek.com>.
- Shute, V. J., Graf, E. A. and Hansen, E. G. (2006). *Designing Adaptive, Diagnostic Math Assessments for Individuals With and Without Visual Disabilities*. Educational Testing Service Research Report #RR-06-01. Downloaded July 2006 from <http://www.ets.org/Media/Research/pdf/RR-06-01.pdf>

Title: Assessing Mathematical Knowledge in a Learning Space: Validity and Reliability

Author: Jean-Claude Falmagne, Cognitive Sciences, UC – Irvine

By design, the assessment of knowledge in a learning space for a scholarly topic, such as beginning algebra, is comprehensive in that the types of problems that can be asked come from a collection encompassing the full curriculum for the topic. The product of an assessment is a knowledge state gathering all the types of problems that the student is capable of solving. Typically, the number of feasible knowledge states is large, on the order of 10^8 . The duration of an assessment is nevertheless tolerable, ranging around 25-35 problems. We report the results of a large scale study (248,502 assessments) investigating whether such an assessment is predictive of the subject's responses to problems that are not part of the assessment. In each assessment, an extra question was asked, the response to which is predictable from the assessed state. The median correlation between predicted and observed correct responses was around .78.

Title: ACED Evaluation

Authors: Valerie Shute (Research Applications & Development, ETS), Eric Hansen (Research Applications & Development, ETS) & Russell Almond (Psychometric Development & Resources, ETS)

This paper will describe the results of an experimental evaluation of ACED: Adaptive Content with Evidence-based Diagnosis. ACED is a computer delivered, diagnostic assessment system for algebra I content. One goal of the research was to examine specific

contributions to learning of various design elements (i.e., adaptivity and diagnostic feedback). Another goal was to test the effectiveness of the system for sighted and visually-disabled students (i.e., blind and low vision). We thus tested the degree to which ACED works—both in terms of being a valid assessment tool and an effective way to support learning with about 300 sighted students. We also tested the system's ability to communicate algebraic content to students with visual disabilities integrating a talking tactile tablet (from Touch Graphics, Inc.) into the ACED design. Results show that the system works well (a) as a valid assessment tool, and (b) in support of learning for students with various abilities.

Title: Relieving Accountability Pressures on Teachers with an Online System that Teaches as it Assesses

Authors: Neil Heffernan (CS, WPI) & Ken Koedinger (HCII, Carnegie Mellon)

NCLB puts teachers in the dilemma of wanting to use formative assessments to be data-driven, but these assessments take away valuable instructional time. We report on our assessment intervention, called ASSISTments, which tutors students for a NCLB test while it assesses. The ASSISTment system is an effort to provide assessment of students' 8th grade math knowledge, while at the same time fine-grained reporting to teachers including NCLB predictions. 18 teachers used the system as part of their normal math classes in 2005-06. Does providing assistance hurt the accuracy of the assessment? Surprisingly, we report, the tutoring actually improves the assessment by tracking how much assistance they need. Furthermore, our fine-grained cognitive models that map each question to a few knowledge components allow us to more accurately predict these scores. We also report on the evidence that students are learning during the tests.

Title: Uncertainty, Prediction and Teacher Feedback using an Online System that Teaches as it Assesses

Author: Brian Junker (Statistics, Carnegie Mellon)

Heffernan & Koedinger's Assistments system has two assessment goals: predicting student performance on end-of-year accountability exams; and helping teachers target next instructional steps. These goals are complicated by two features of the system: assessment is ongoing throughout the school year as student proficiency develops; and the Assistments system itself changes student proficiency. Nevertheless, prediction using simple student models and a number of "assistance metrics" (summaries of hint-seeking behavior, time spent on questions correctly vs. incorrectly answered, etc.) can be almost optimally effective at predicting end-of-year exam scores; more complex student models provide marginal improvement. Statistical uncertainty in teacher feedback reports based on more-detailed student models is sometimes surprisingly low, but even in cases where the per-student uncertainty is high, reports aggregated over groups of students can be quite reliable. These ideas will be considered both for the Assistments system and in the broader context of online assessment and learning systems.